Searching superhard C and $B_xC_yN_z$ via full-space inverse design approach

Kaiwei Feng [1], Guanjian Cheng [2], and Wan-Jian Yin [1,3,*]

[1] College of Energy, Soochow Institute for Energy and Materials InnovationS (SIEMIS), and Jiangsu Provincial Key Laboratory for Advanced Carbon Materials and Wearable Energy Technologies, Soochow University, Suzhou 215006, China

[2] Key Laboratory for Computational Physical Sciences (MOE), Institute of Computational Physical Sciences, Fudan University, Shanghai 200438, China

[3] Hefei National Laboratory, Hefei 230088, China

(Received 31 August 2025; accepted 25 November 2025; published 11 December 2025)

The inverse design of superhard materials is often hindered by the vastness of the configurational space. In this work, we employ our recently developed D^{3}REAM framework—an approach that integrates machine-learning potentials with optimization strategies—to enable automated and efficient discovery of superhard materials. Using D^{3}REAM, we identified carbon allotropes, $C_8(C2/m)$ and $C_8(I2_12_12_1)$, with bulk moduli approaching 400 GPa. Furthermore, we found four carbon allotropes, $C_4(I2_12_12_1)$, $C_8(I2_12_12_1)$, $C_8(P\bar{1})$, and $C_8(C2/m)$, exhibiting uniaxial elastic moduli surpassing that of diamond. Extending the search to the complex B-C-N system, active learning revealed superhard phases $B_2C_5N(Pmm2)$, $B_2C_3N(Cm)$, and $B_2C_{16}N_4(P1)$. Analysis of the B-C-N search space further indicates that boron-nitrogen incorporation up to 40% enables the design of superhard materials with enriched structural and mechanical diversity.

DOI: 10.1103/8qyx-sqtn

I. INTRODUCTION

Materials with a Vickers hardness above 40 GPa are classified as superhard materials. Owing to their exceptional mechanical strength, they are widely employed in industrial applications such as cutting, drilling, boring, turning, and grinding [1–5]. Diamond, the hardest material known to date, possesses a Vickers hardness of approximately 90–100 GPa and an ultrahigh Young's modulus, rendering it highly resistant to wear and virtually undeformable under applied stress [6]. Nevertheless, its widespread use in manufacturing is limited by several drawbacks, including intrinsic brittleness, poor low-temperature plasticity, processing challenges, and high production cost [6,7]. Consequently, the search for new superhard materials with superior overall performance has become a critical research focus.

The chemical bonds among light elements (B, C, N, O, etc.) are typically strong covalent bonds, which allow them to form dense and stable atomic networks [8]. These strong interactions facilitate the formation of superhard materials with complex structures. As a result, the search for superhard materials within light-element systems has drawn significant attention and yielded several notable discoveries, such as C_3N_4 [9,10], cubic boron nitride (c-BN) [11,12], BC_5 [13–15], and boron-rich $B_{50}C_2$ [16–18]. Nevertheless, due to the vast phase space across different systems and compositional combinations, many regions of this landscape remain unexplored, leaving substantial opportunities for discovering new superhard materials.

Earlier studies mainly relied on first-principles density functional theory (DFT) simulations, where mechanical properties were evaluated using the energy-strain method [19,20]. However, due to the high computational cost, such studies typically required prior knowledge of atomic composition, stoichiometric ratios, or crystal structures to reduce the search space. To overcome these limitations, machine learning (ML) approaches for accelerating inverse materials design have gained widespread attention [21–23]. Conventional ML algorithms, such as random forests, have achieved notable successes in the search for superhard materials. For example, $BC_{10}N$, $B_4C_5N_3$, and B_2C_3N were identified in the B-C-N system [24], while $B_5N_3O_3$, $B_6N_4O_3$, $B_7N_5O_3$, and $B_9N_7O_3$ were discovered in the B-N-O system [25].

Conventional supervised-learning pipelines—most notably hand-crafted feature engineering followed by support-vector classification—demand labor-intensive descriptor construction, scale poorly to high-dimensional spaces, and remain highly vulnerable to noise [26]. Furthermore, optimization algorithms generally confine inverse design to local regions of the search space defined by prior assumptions. To address these issues, we recently developed a global optimization framework, implemented in the D^{3}REAM (data-driven design and rapid exploration of advanced materials) software [27]. This method enables free and comprehensive exploration of the materials space without requiring prior information on atomic composition, stoichiometry, or crystal structure. By embedding machine-learning-based force field models [28,29], D^{3}REAM can rapidly and accurately predict crystal structures according to target properties, providing a powerful new paradigm for superhard material discovery.

In this article, we employed the D^{3}REAM software to search for superhard C and $B_xC_yN_z$ compounds using bulk

*Contact author: wjyin@suda.edu.cn

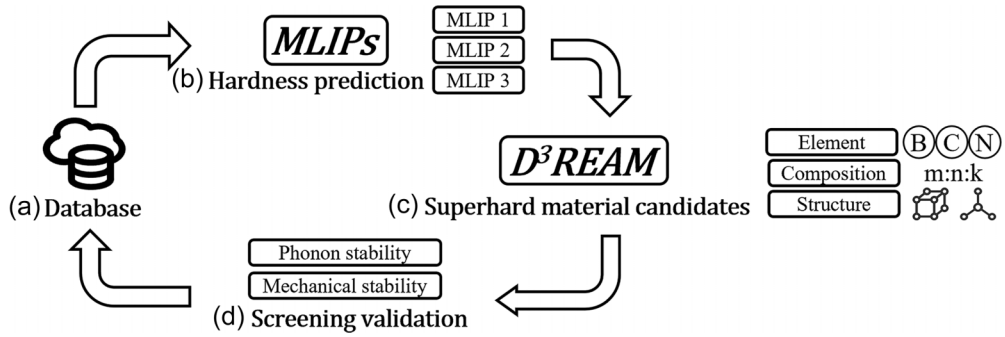

FIG. 1. Data-driven computational framework for discovering superhard materials. (a) Data set for machine learning potential function training. (b) Hardness prediction by setting different random seeds. (c) Searching candidates for superhard materials by using D^{3}REAM crystal software embedded with machine learning potential function. (d) Screening out inaccurate structures, verifying calculation by VASP based on first principles, and putting inaccurate structures into data sets for retraining.

modulus and density as the target properties, following a full-space inverse design strategy. The predicted results were further validated by DFT calculations of bulk modulus, shear modulus, Young's modulus, and Vickers hardness of the predicted results. Through this approach, we identified severalfound stable superhard materials that are both thermodynamically and dynamically stable. Specifically, we two new carbon allotropes, $C_8(C2/m)$ and $C_8(I2_12_12_1)$, as well as $B_2C_5N(Pmm2)$, $B_2C_3N(Cm)$, and $B_2C_{16}N_4(P1)$ in the B-C-N system. Notably, some of these structures exhibit strong elastic anisotropy and possess maximum elastic moduli exceeding that of diamond, as detailed in of Sec. III C. The overall computational workflow is summarized in Fig. 1 and discussed in detail in Sec. II.

II. METHODS

A. Machine learning force field

For the carbon system, we used *ab initio* random structure searching (AIRSS) based on the random search algorithm to generate the training dataset [30]. Specifically, we applied a pressure of 10 GPa to promote atomic rearrangement and enhance structural diversity, while constraining the maximum number of atoms per unit cell to 24 for random structure sampling. From 11 370 candidate structures, duplicates were removed, yielding 8270 unique structures, which were then optimized using the Vienna *ab initio* simulation package (VASP) [31]. For each converged calculation, we uniformly extracted the structural, energy, force, and virial data at ten points along the optimization path, ultimately obtaining 61 781 energy-point entries.

For the B-C-N system, which is more complex than the pure carbon system, it is difficult to directly construct a reliable machine-learning force-field model using simple random sampling. To address this, we adopted an active learning strategy (Fig. 1), iteratively improving the dataset through multiple rounds of model training, structure searching, and retraining. Concretely, three machine-learning force-field models were trained with identical hyperparameters but different random seeds. For candidate structures where the predicted hardness exhibited a standard deviation greater than 1 GPa among the three models, we performed first-principles validation. As in the carbon case, ten energy points were uniformly extracted during the optimization process for each converged calculation, and data corresponding to large deviations in model predictions were added to the dataset for further refinement. The initial dataset in this case was derived by introducing random perturbations to both lattice parameters and atomic positions of structures taken from the OQMD database, followed by static self-consistent relaxation. This procedure yielded 8792 energy-point entries.

For training the machine-learning force-field model, we employed the MACE framework [32], a recently developed architecture based on atomic cluster expansion (ACE) that integrates E(3)-equivariant message passing with efficient multibody messaging [33–35]. By constructing high-order features through a hierarchical body-order expansion, MACE avoids the exponential cost growth typically associated with feature order [36]. This enables the development of force-field models that can accurately describe complex interatomic interactions and be efficiently computed in parallel.

For each model training, the dataset was divided into training, validation, and test sets in an 80:10:10 ratio. The model was configured as ScaleShiftMACE with a batch size of 100, a maximum of 1600 training epochs, and an early-stopping patience of 70. Random weight averaging was applied starting from the 1500th epoch. Prior to random weight averaging, the loss-function weights for energy, force, and virial were set to 10, 5, and 0.1, respectively. During the averaging process, these weights were adjusted to 1, 1, and 0.1. All training was carried out on CUDA-enabled GPUs.

B. Superhard materials crystal structure prediction

Currently, crystal structure prediction (CSP) primarily aims to identify the most thermodynamically stable crystal structure, i.e., the atomic arrangement with the lowest energy, without relying on prior experimental information. Mathematically, this corresponds to solving an optimization problem of locating local minima on a high-dimensional potential energy surface [37]. In contrast, for the problem addressed in this study, the optimization target is the Vickers hardness. Specifically, we seek the crystal structure—defined by its chemical composition and atomic arrangement—that maximizes Vickers hardness within a given system.

Here, we use D^{3}REAM developed by Cheng *et al.* to search for superhard materials. It explores the phase space without prior knowledge of stoichiometry, crystal structure, or atomic composition, and supports mainstream ML force-field interfaces for data-driven materials discovery. For optimization, we employ the Fast nondominated sorting genetic al gorithm (NSGA-II) to balance hardness and density. NSGA-II uses congestion distance to maintain diversity and prevent premature clustering [38,39], while its elite retention strategy preserves high-quality individuals across generations. Together, these features improve convergence efficiency and enable the algorithm to identify superior solutions more rapidly.

Specifically, we embed the trained MACE force field into D^{3}REAM for crystal structure prediction. First, the crystal structures suggested by NSGA-II [40], are preliminarily optimized with the LJ potential, and unreasonable ones are removed by interatomic distance screening. The ML force field then predicts Vickers hardness, with density and hardness as optimization targets for the next NSGA-II iteration. Introducing density both avoids costly hardness predictions on low-density structures and, with an upper bound, prevents unphysical short interatomic distances. In addition, the ML force field predicts XRD spectra [41], and cosine similarity is used to ensure newly proposed structures differ from those already found, thus preventing trapping in local minima.

C. First principles calculations

DFT calculations were performed using VASP with input files generated by Pymatgen [42].

Structural optimization and self-consistent calculations employed default convergence criteria, with a plane-wave cutoff of 520 eV. The PBE functionals were used, with spin neglected. Elastic constants were obtained by applying five strains (0, ±0.006, ±0.012) via VASPKIT [43], and bulk/shear moduli were derived using the Voigt-Reuss-Hill scheme. Vickers hardness was then estimated using Chen's empirical formula [44–47]:

$$H_v = 2(k^2G)^{0.585} - 3.0,$$

where Pugh's ratio k(= G/K).

To ensure reliability, quadratic optimization and SCF calculations were followed by band-structure analysis with MPNonSCFSet in Pymatgen. Phonon dispersion was calculated with PHONOPY [48], using DFPT on a $2 \times 2 \times 2$ supercell. For hardness, band, and phonon calculations, stricter convergence thresholds for energy and force were adopted.

D. Structure similarity comparison

To ensure independence of the initial training set and novelty of structures generated during prediction, we compared structural similarities using the Niggli algorithm implemented in Pymatgen's StructureMatcher class [49]. Fractional length and site tolerances were set to 0.1, and the angle tolerance to 1°. Structures below these thresholds were regarded as identical.

TABLE I. MAE in training MACE potential function from scratch. The training average absolute error of the three different random seeds on the test set of carbon. From the test, their performance is similar and their accuracy is high.

Model	MAE_{energy} (meV/atom)	MAE_{force} (meV/Å)	MAE_{virals} (meV)
1	12.06	69.53	774.17
2	12.95	73.29	781.32
3	12.07	70.23	765.99

III. RESULTS AND DISCUSSION

A. Carbon system

Figures 2(b)–2(d) show the distribution histograms of energy, force, stress, and density for the dataset used to train the machine learning force field in the carbon system. The dataset contains 61 781 configurations. To evaluate model performance, we compared three state-of-the-art ML force fields: MACE, which employs advanced-order isovariable message passing [50]; CHGNet, which efficiently represents multibody interactions [51]; and NEP, specifically designed for carbon systems [21]. As shown in Figs. 3(a)–3(c), pretrained models of these potentials were tested, and while MACE exhibited the strongest predictive ability, its accuracy remained insufficient. We therefore retrained MACE on our dataset. After *de novo* training, the accuracy of the MACE model improved markedly, confirming its effectiveness for crystal structure prediction [Fig. 3(d)].

To better capture atomic interactions in carbon, we trained three models using different random seeds under identical parameters. As summarized in Table I, all models achieved consistently high predictive accuracy on the test set, demonstrating both robustness and strong generalization. Benchmarks for prediction errors of these models are shown in Figs. S1 and S2 of the Supplemental Material [52]. Building on this trained model, we further computed the elastic constant matrix via strain-energy relations, from which bulk modulus (K) and shear modulus (G) were derived. Finally, Vickers hardness was estimated using Chen's empirical formula, validating the model's reliability for describing interatomic interactions in carbon systems.

We investigated the carbon system using the D^{3}REAM platform, which integrates a machine learning force-field modeling API. The search was limited to candidate structures containing no more than 24 atoms, with duplicates removed through structural similarity screening. From the resulting dataset, we retained all structures with protocell sizes below eight atoms and additionally selected the top 100 structures with the highest predicted hardness among those exceeding this threshold. The elastic constants of these selected structures were subsequently validated through first-principles calculations using VASP.

We first excluded structures with negative eigenvalues in the elastic constant matrix to ensure mechanical stability. We also removed materials with Pugh's ratio $k < 0.25$, which indicates extremely low hardness, and those with $k > 4.0$, corresponding to unrealistically high hardness values

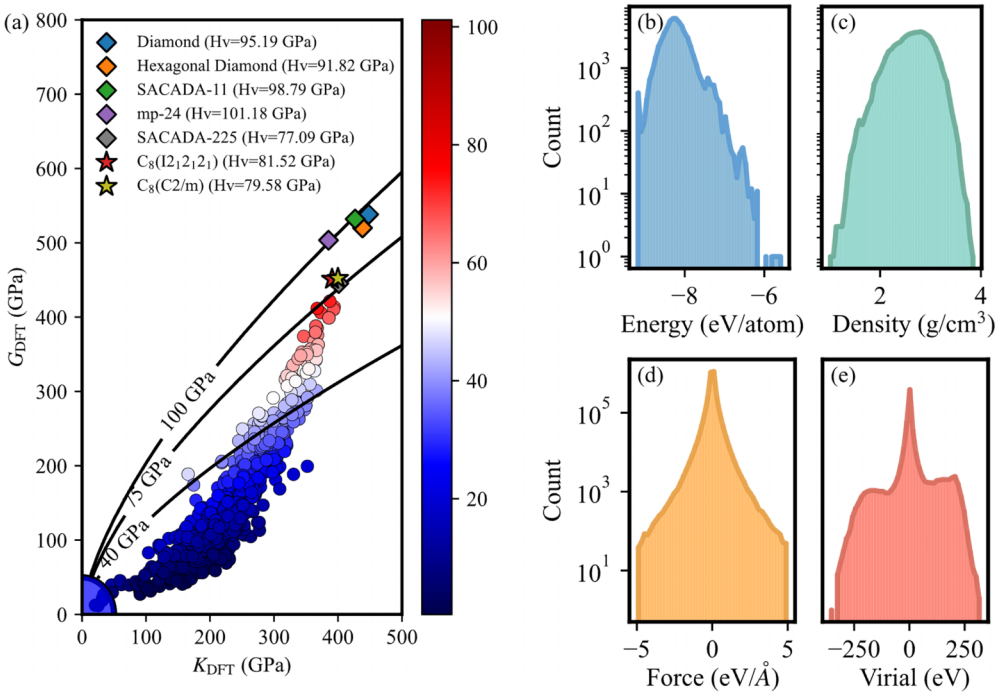

FIG. 2. Search for carbon system superhard materials. (a) False color intensity represents Vickers hardness calculated using K and G as inputs. The solid lines represent the Vickers hardness contour profile calculated based on Chen's empirical model, which can help quickly locate compounds with superhard materials. The undiscovered superhard materials with a hardness of more than 75 GPa that were found are represented by ☆, and the existing superhard materials in MP and SACADA databases are represented by △. (b) Data distribution of energy. (c) Data distribution of force. (d) Data distribution of virial. (e) Data distribution of and density.

(H > 200 GPa) typically observed only under high pressure. Figure 2(a) presents the scatter plot of bulk versus shear modulus for the searched structures, with false-color depth representing Vickers hardness estimated by Chen's empirical formula. Notably, results in the low-bulk-modulus region appear unreliable, reflecting the known limitations of this

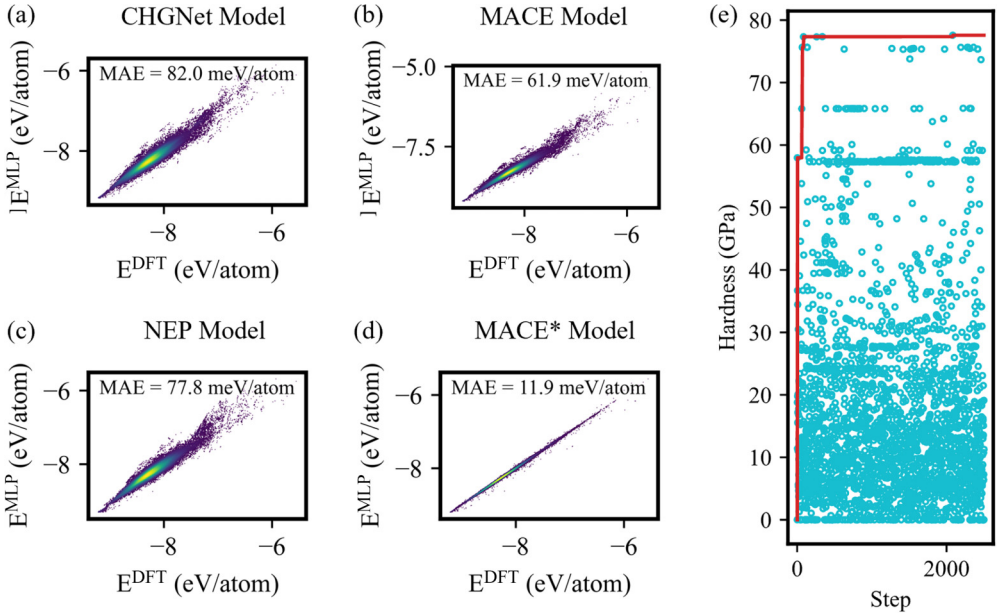

FIG. 3. Comparison of the prediction capabilities of machine learning potential function prediction for the sampled carbon material datasets. (a) CHGNet pretrained model, (b) MACE pretrained model, (c) NEP universal model for Carbon system, and (d) MACE model trained from scratch using dataset. Through kernel density estimates, the color from cool to warm represents the density of scattered points at this point. (e) Search for trajectories of superhard materials using D^{3}REAM software embedded in the refractory training MACE potential function.

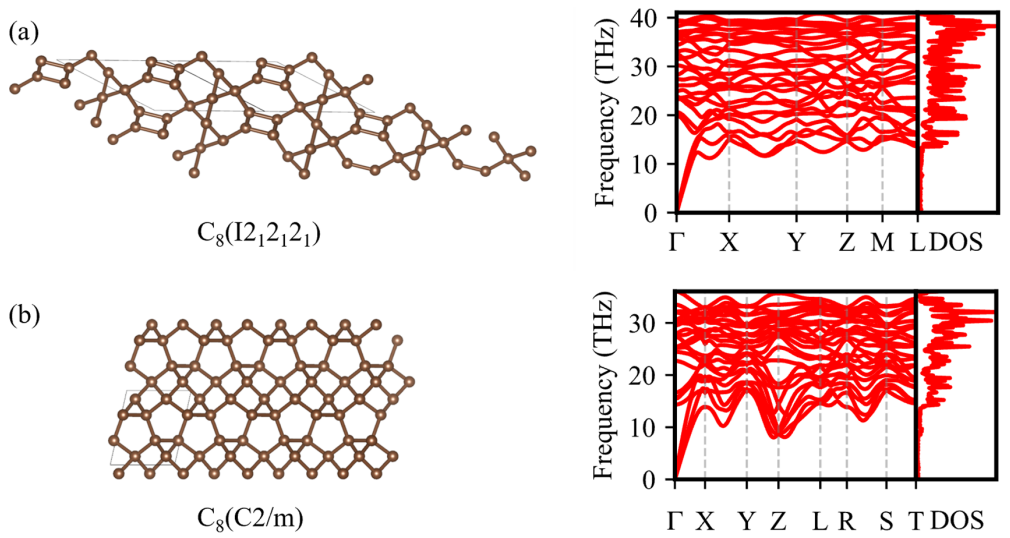

FIG. 4. Optimization algorithm and density functional theory calculation for carbon system. Theoretical crystal structure (left) and phonon dispersion relation (right) of (a) $C_8(I2_12_12_1)$ and (b) $C_8(C2/m)$. Both show dynamic stability (phonon spectra have no imaginary frequencies).

empirical relation. Contour lines for hardness values of 40, 75, and 100 GPa are also shown. For candidate superhard materials ($H > 75$ GPa), we crosschecked against the MP and SACADA carbon databases, marking previously reported structures with ☆ and newly identified ones with □.

For the superhard phases $C_8(I2_12_12_1)$ and $C_8(C2/m)$, both exhibiting Vickers hardness above 75 GPa, we obtained bulk moduli of 390.66 GPa and 399.90 GPa, and shear moduli of 451.15 GPa and 452.23 GPa, respectively. Phonon spectra were calculated for both structures, as shown in Figs. 4(a) and 4(b). The absence of imaginary modes and the broad stable frequency ranges confirm that these materials are dynamically stable. Their lattice parameters, densities, Vickers hardness, band gaps, and average atomic energies are summarized in Table II.

Since bulk and shear moduli are often anisotropic, our reported values are directional averages. We further analyzed anisotropy using the elastic constant matrices [53]. As shown in Figs. 5(a)–5(d), the maximum bulk, shear, and Young's moduli of $C_8(C2/m)$ occur along the (001) surface, reaching 453.71 GPa, 490.52 GPa, and 1126.66 GPa, respectively. Notably, the maximum bulk modulus of this phase exceeds that of diamond, while its maximum Vickers hardness (89.42 GPa) approaches diamond. Based on first-principles verification, we identified superhard materials with minimum hardness above 40 GPa and maximum elastic moduli exceeding diamond. As summarized in Table III, four carbon allotropes—$C_4(I2_12_12_1)$, $C_8(I2_12_12_1)$, $C_8(P\bar{1})$, and $C_8(C2/m)$—exhibit bulk modulus values surpassing diamond in certain directions. For these materials, see Fig. S4 of the Supplemental Material [52] for their structure and phonon dispersion relations, and Table S1 of the Supplemental Material [52] for information on lattice parameters. This indicates that, within carbon systems, bulk modulus is the hardness index most likely to exceed diamond. Such exceptional stiffness provides superior hardness, wear resistance, and dimensional stability, making these materials promising for nanocomposites, protective coatings, and precision mechanical components [54], as well as for designing extreme micro- and nanoscale structures.

B. Boron-carbon-nitrogen system

We applied our approach to search for superhard materials in the boron-carbon-nitrogen ($B_xC_yN_z$) system. Unlike elemental carbon, this system allows variation in both unit-cell size and composition, providing a stringent test for our optimization algorithm. Due to the complex potential energy surface and limited available data, we used active learning to train ML force-field models. Three models were trained under the MACE framework with different random seeds for energy, force, and virial predictions. Figures 6(a)–6(c) show the MAE of these models across epochs, while Fig. 6(d) tracks the training data size. After six rounds of active learning, the final test-set MAEs were 18.01 meV/atom (energy), 109.70 meV/Å (force), and 874.21 meV (virial), demonstrating effective description of interatomic interactions. Benchmark of the error

TABLE II. Structural and performance parameters based on DFT theoretical calculation. Among them, a (Å), b (Å), c (Å), α, β, and γ represent the lattice constants of the structure, ρ represents the density of the structure (g/cm^3), H_v is the Vickers hardness of the structure (GPa), Gap represents the energy band gap of the structure (eV), and E represents the average atomic energy of the structure (eV/atom).

Formula	a	b	c	α	β	γ	ρ	H_v	Gap	E
$C_8(I2_12_12_1)$	6.27	6.27	6.27	155.76	153.86	36.36	3.51	81.52	2.62	−7.83
$C_8(C2/m)$	3.46	3.46	4.18	78.54	101.46	86.17	3.35	79.58	3.84	−8.89

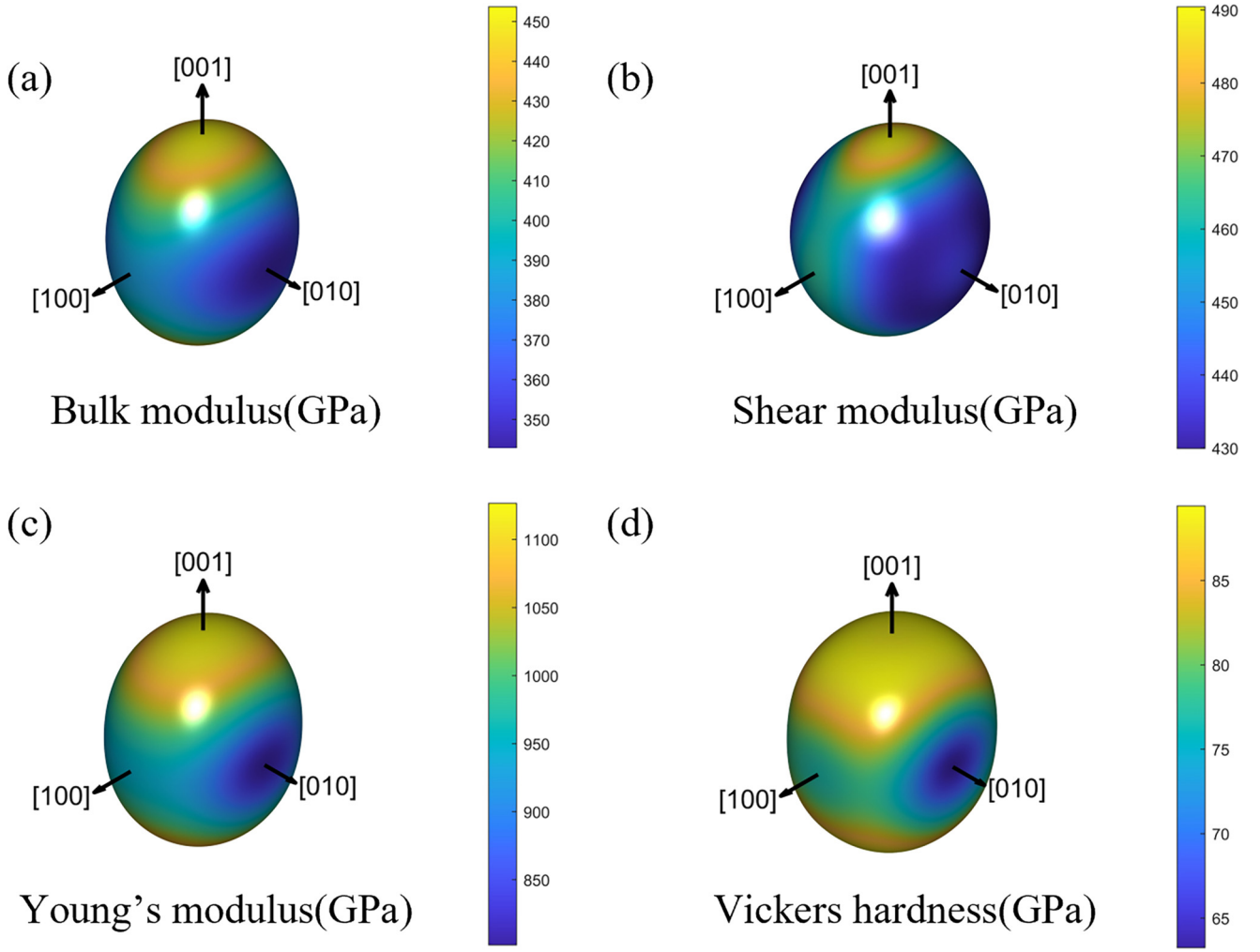

FIG. 5. Anisotropic analysis of elastic constants. Anisotropic distribution of (a) bulk modulus, (b) shear modulus, (c) Young's modulus, and (d) Vickers hardness. False color from dark to light indicates that the elastic modulus increases from small to large. Among them, the maximum values of bulk modulus, shear modulus, and Young's modulus all appear on the (001) surface.

of the final model is shown in Fig. S3 of the Supplemental Material [52].

Using these models, we searched structures with up to 20 atoms of B, C, and N. After removing duplicates by structural similarity, we retained the top 100 candidates ranked by bulk, shear, Young's moduli, and Vickers hardness in each round, and validated them with first-principles elastic constant calculations in VASP. Results were further screened for mechanical stability and Pugh's ratio ($0.25 < k < 4$). Figure 7(a) shows the bulk vs shear modulus distribution, while ternary phase diagrams in Figs. 7(a)–7(e) reveal that high-carbon regions tend to yield larger elastic moduli, consistent with the strength of C-C bonds. Interestingly, superhard materials also cluster near 60% carbon content, suggesting that mixing $\sim$ 40% B and N with carbon is a promising strategy for designing superhard materials with diverse properties.

As shown in Figs. 8(a)–8(c), three compounds—B_2C_5N(Pmm2), B_2C_3N(Cm), and $B_2C_{16}N_4$(P1)—exhibit

TABLE III. Superhard materials that exceed diamond in maximum direction. Among them, K_{max}, G_{max}, E_{max}, and H_{max} represent the maximum values of bulk modulus, shear modulus, Young's modulus, and Vickers hardness; Modulus represents the modulus beyond diamond; A_K, A_G, A_E, and A_H represent the anisotropy index of the structure in bulk modulus, shear modulus, Young's modulu,s and Vickers hardness, and their values are equal to the maximum value of the elastic modulus of the structure divided by the minimum value.

Structure	Modulus	K_{max}	A_K	G_{max}	A_G	H_{max}	A_H	E_{max}	A_E
$C_4(I2_12_12_1)$	K	449.44	1.77	560.36	1.86	100.27	2.07	1156.93	1.98
$C_8(I2_12_12_1)$	K	453.71	1.32	521.23	1.36	89.42	1.41	1126.66	1.40
$C_8(P\bar{1})$	K,H	513.29	2.04	569.81	2.29	122.50	3.00	1224.32	2.09
$C_8(C2/m)$	K	468.45	1.70	489.97	1.58	96.48	1.77	1194.96	1.84

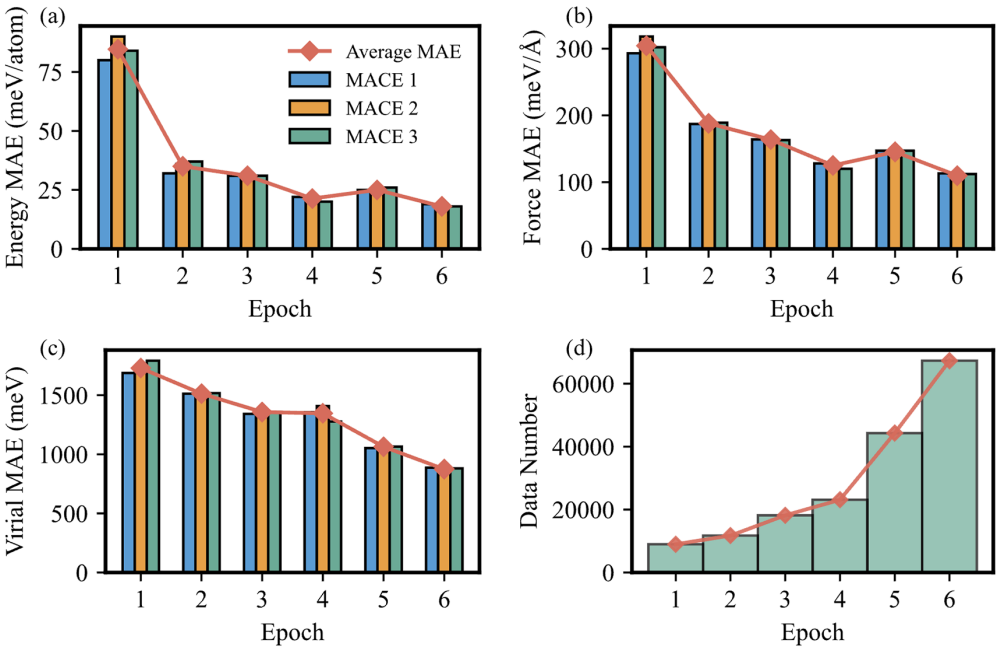

FIG. 6. Active learning training results. After six rounds of active learning, the bar graph distribution of the MAE predicted by three machine learning models based on the MACE framework with random seeds of 1, 3, and 7 for (a) energy, (b) force, and (c) virial. The broken line represents the average MAE of the prediction results of the three machine learning models. (d) is a bar graph of the training data during the six rounds of active learning.

Vickers hardness above 70 GPa, with bulk moduli of 356.88, 341.14, and 399.05 GPa, and shear moduli of 416.19, 380.77, and 421.74 GPa, respectively. Phonon spectra confirm their dynamic stability.

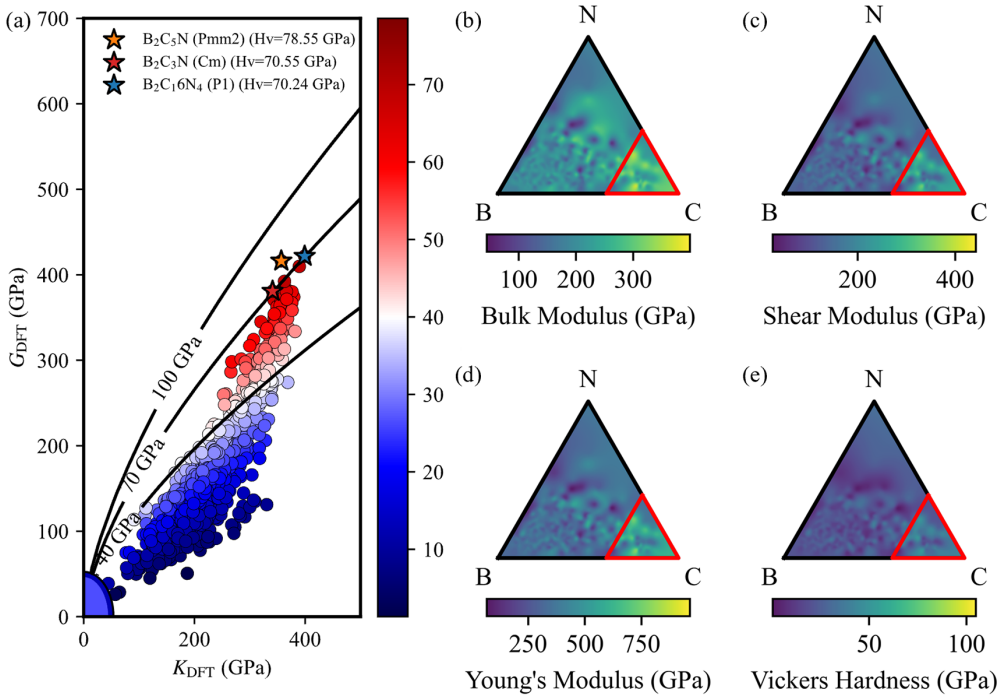

FIG. 7. Search results for superhard materials in the B-C-N system. (a) Scatter plot of the bulk modulus and shear modulus distribution of the search results. The false color intensity indicates the size of the Vickers hardness. For ease of observation, we also drew the Vickers hardness contours of 40, 70, and 100 GPa, and used ☆ to indicate the superhard materials with Vickers hardness exceeding 70 GPa that have never been found. (b) Ternary phase diagram distribution of the search results on bulk modulus, (c) shear modulus, (d) Young's modulus, and (e) Vickers hardness. The high modulus area mainly appears in the area with a carbon content of more than 60%.

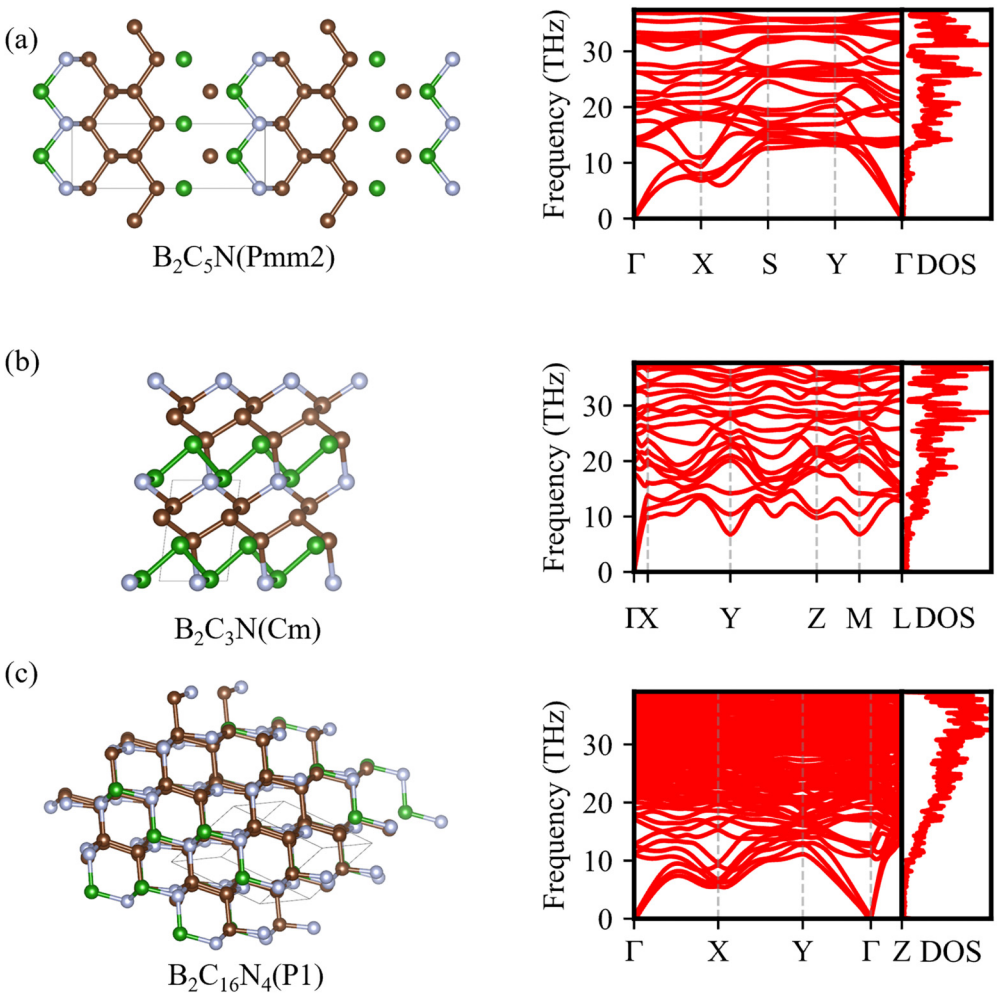

FIG. 8. Optimization algorithm and density functional theory calculations for the boron-carbon-nitrogen system. The structures (left) and phonon dispersion relations (right) of (a) B_2C_5N(Pmm2), (b) B_2C_3N(Cm), and (c) $B_2C_{16}N_4$(P1) are all dynamically stable (no imaginary frequencies in the phonon spectra).

Formation energies were also calculated as the difference between the total energy of each structure and the weighted sum of reference energies for β-B, graphite, and nitrogen, as

$$\Delta E = \frac{E_{\text{total}}(B_xC_yN_z) - xE(B) - yE(C) - zE(N)}{x+y+z}.$$

As listed in Table IV, these phases, B_2C_5N(Pmm2), B_2C_3N(Cm), and $B_2C_{16}N_4$(P1), crystallize in orthorhombic, hexagonal, and monoclinic systems. Among them, B_2C_5N(Pmm2) shows the lowest formation energy (∼100 meV/atom), indicating possible experimental synthesizability. Moreover, B_2C_5N(Pmm2) and B_2C_3N(Cm) exhibit zero band gaps, suggesting conductivity. Such materials—combining extreme hardness with electrical conductivity—hold great potential for applications in contact electrodes, wear-resistant conductive parts, high-temperature electrodes, and integrated mechanical-electronic systems [55].

IV. CONCLUSIONS

In this work, we developed a machine learning force-field-based approach for inverse design of materials and

TABLE IV. Structural and performance parameters based on DFT theoretical calculation. Among them, a (Å), b (Å), c (Å), α, β, and γ represent the lattice constants of the structure; ρ represents the density of the structure (g/cm^3); H_v is the Vickers hardness of the structure (GPa); Gap represents the energy band gap of the structure (eV); and ΔE represents the average atomic formation energy of the structure based on the most stable element (meV/atom).

Formula	a	b	c	α	β	γ	ρ	H_v	Gap	ΔE
B_2C_5N	2.53	2.53	7.57	90.00	90.00	90.00	3.27	78.55	0.00	116.18
B_2C_3N	4.54	4.54	3.86	65.19	114.81	148.93	3.48	70.55	0.00	162.60
$B_2C_{16}N_4$	3.60	4.38	9.08	94.24	100.62	112.09	3.21	70.24	2.65	188.50

successfully discovered several new superhard phases in C and B-C-N systems. The method greatly reduces computational cost and enables comprehensive data-driven exploration. Some of the predicted materials exhibit exceptional mechanical properties, with elastic moduli in certain directions surpassing diamond, highlighting the potential of ML-assisted discovery in advanced material design. A major challenge remains in translating theoretical predictions into experimentally synthesizable materials. Future research should focus on exploring synthesis routes and practical applications of these candidates. Meanwhile, improving the accuracy and reliability of ML potentials—through larger training datasets, optimized model architectures, and integration of advanced ML techniques—will enhance predictive power, reduce experimental trial-and-error, and further accelerate discovery. Beyond superhard materials, this ML-driven design paradigm can be extended to other advanced systems such as superconductors, magnetic materials, and energy storage compounds, offering a powerful framework for materials science.

ACKNOWLEDGMENTS

This work was supported by the Quantum Science and Technology National Science and Technology Major Project (Grant No. 2024ZD0300102), National Natural Science Foundation of China (Grant No. 12504263).

DATA AVAILABILITY

All data supporting the findings of this study are openly available at [56].

[1] A. Friedrich, B. Winkler, E. A. Juarez-Arellano, and L. Bayarjargal, Synthesis of binary transition metal nitrides, carbides and borides from the elements in the laser heated diamond anvil cell and their structure-property relations, Materials **4**, 1648 (2011).

[2] A. G. Kvashnin, Z. Allahyari, and A. R. Oganov, Computational discovery of hard and superhard materials, J. Appl. Phys. **126**, 040901 (2019).

[3] Z. Zhao, B. Xu, and Y. Tian, Recent advances in superhard materials, Annu. Rev. Mater. Res. **46**, 383 (2016).

[4] M. T. Yeung, R. Mohammadi, and R. B. Kaner, Ultraincompressible, superhard materials, Annu. Rev. Mater. Res. **46**, 465 (2016).

[5] Y. Le Godec, A. Courac, and V. L. Solozhenko, High-pressure synthesis of superhard and ultrahard materials, J. Appl. Phys. **126**, 151102 (2019).

[6] J.-J. Zhao, S. Scandolo, J. Kohanoff, G. L. Chiarotti, and E. Tosatti, Elasticity and mechanical instabilities of diamond at megabar stresses: Implications for diamond-anvil-cell research, Appl. Phys. Lett. **75**, 487 (1999).

[7] F. Zhao, Y. He, B. Huang, T. Zhang, and H. Zhu, A review of diamond materials and applications in power semiconductor devices, Materials **17**, 3437 (2024).

[8] O. O. Kurakevych, Superhard phases of simple substances and binary compounds of the B-C-N-O system: From diamond to the latest results (a Review), J. Superhard Mater. **31**, 139 (2009).

[9] Z. Yan, H. Chuanzhen, and L. Hanlian, Sintering and mechanics properties of C_3N_4 based ceramic tool materials, China Mech. Eng. **34**, 352 (2023).

[10] J. Yang, X. Zhang, C. Xie, J. Long, Y. Wang, L. Wei, and X. Yang, Preparation of g-C_3N_4 with high specific surface area and photocatalytic stability, J. Electron. Mater. **50**, 1067 (2021).

[11] S. N. Monteiro, A. L. D. Skury, M. G. De Azevedo, and G. S. Bobrovnitchii, Cubic boron nitride competing with diamond as a superhard engineering material–an overview, J. Mater. Res. Technol. **2**, 68 (2013).

[12] E. Benko, J. S. Stanisław, B. Królicka, A. Wyczesany, and T. L. Barr, cBN–TiN, cBN–TiC composites: Chemical equilibria, microstructure and hardness mechanical investigations, Diam. Relat. Mater. **8**, 1838 (1999).

[13] Q. Li, H. Wang, Y. Tian, Y. Xia, T. Cui, J. He, Y. Ma, and G. Zou, Superhard and superconducting structures of BC_5, J. Appl. Phys. **108**, 023507 (2010).

[14] V. L. Solozhenko, O. O. Kurakevych, D. Andrault, Y. Le Godec, and M. Mezouar, Ultimate metastable solubility of boron in diamond: Synthesis of superhard diamondlike BC_5, Phys. Rev. Lett. **102**, 015506 (2009).

[15] P. A. Baker, S. A. Catledge, S. B. Harris, K. J. Ham, W.-C. Chen, C.-C. Chen, and Y. K. Vohra, Computational predictions and microwave plasma synthesis of superhard boron-carbon materials, Materials **11**, 1279 (2018).

[16] P. A. Baker, W.-C. Chen, C.-C. Chen, S. A. Catledge, and Y. K. Vohra, First-principles predictions and synthesis of $B_{50}C_2$ by chemical vapor deposition, Sci. Rep. **10**, 4454 (2020).

[17] K. Chakrabarty, W.-C. Chen, P. A. Baker, V. M. Vijayan, C.-C. Chen, and S. A. Catledge, Superhard boron-rich boron carbide with controlled degree of crystallinity, Materials **13**, 3622 (2020).

[18] N. Uemura, K. Shirai, H. Eckert, and J. Kunstmann, Structure, nonstoichiometry, and geometrical frustration of α -tetragonal boron, Phys. Rev. B **93**, 104101 (2016).

[19] M. Zhang, H. Liu, Q. Li, B. Gao, Y. Wang, H. Li, C. Chen, and Y. Ma, Superhard BC_3 in cubic diamond structure, Phys. Rev. Lett. **114**, 015502 (2015).

[20] S. Zhang, J. He, Z. Zhao, D. Yu, and Y. Tian, Discovery of superhard materials via CALYPSO methodology*, Chin. Phys. B **28**, 106104 (2019).

[21] K. Song *et al.*, General-purpose machine-learned potential for 16 elemental metals and their alloys, Nat. Commun. **15**, 10208 (2024).

[22] J. Behler and G. Csányi, Machine learning potentials for extended systems: A perspective, Eur. Phys. J. B **94**, 142 (2021).

[23] C.-N. Li, H.-P. Liang, B.-Q. Zhao, S.-H. Wei, and X. Zhang, Machine learning assisted crystal structure prediction made simple, J. Mater. Inform. **4**, 15 (2024).

[24] W.-C. Chen, J. N. Schmidt, D. Yan, Y. K. Vohra, and C.-C. Chen, Machine learning and evolutionary prediction of superhard B-C-N compounds, npj Comput. Mater. **7**, 114 (2021).

[25] W.-C. Chen, Y. K. Vohra, and C.-C. Chen, Discovering superhard B–N–O compounds by iterative machine learning

and evolutionary structure predictions, ACS Omega **7**, 21035 (2022).
[26] A. A. Wani, Advancing material stability prediction: Leveraging machine learning and high-dimensional data for improved accuracy, Mater. Sci. Appl. **16**, 79 (2025).
[27] G. Cheng, X.-G. Gong, and W.-J. Yin, Crystal structure prediction by combining graph network and optimization algorithm, Nat. Commun. **13**, 1492 (2022).
[28] G. Cheng, X.-G. Gong, and W.-J. Yin, An approach for full space inverse materials design by combining universal machine learning potential, universal property model, and optimization algorithm, Sci. Bull. **69**, 3066 (2024).
[29] X. Wang, Z. Wang, P. Gao, C. Zhang, J. Lv, H. Wang, H. Liu, Y. Wang, and Y. Ma, Data-driven prediction of complex crystal structures of dense lithium, Nat. Commun. **14**, 2924 (2023).
[30] C. J. Pickard and R. J. Needs, *Ab initio* random structure searching, J. Phys. Condens. Matter **23**, 053201 (2011).
[31] G. Kresse and J. Furthmüller, Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996).
[32] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csanyi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, in *Advances in Neural Information Processing Systems*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Vol. 35 (Curran Associates, Inc., 2022), pp. 11423–11436.
[33] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, Phys. Rev. B **99**, 014104 (2019).
[34] V. G. Satorras, E. Hoogeboom, and M. Welling, E(n) equivariant graph neural networks, in *Proceedings of the 38th International Conference on Machine Learning*, edited by M. Meila and T. Zhang, Vol. 139 (PMLR, 2021), pp. 9323–9332.
[35] N. Ran, L. Yin, W. Qiu, and J. Liu, Recent advances in machine learning interatomic potentials for cross-scale computational simulation of materials, Sci. China Mater. **67**, 1082 (2024).
[36] R. Kondor and S. Trivedi, On the generalization of equivariance and convolution in neural networks to the action of compact groups, in *Proceedings of the 35th International Conference on Machine Learning*, edited by J. Dy and A. Krause, Vol. 80 (PMLR, 2018), pp. 2747–2755.
[37] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SchNet - A deep learning architecture for molecules and materials, J. Chem. Phys. **148**, 241722 (2018).
[38] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, Scalable multi-objective optimization test problems, in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02* (Cat. No.02TH8600), Vol. 1 (IEEE, Honolulu, HI, USA, 2002), pp. 825–830.
[39] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. **6**, 182 (2002).
[40] J. E. Jones, On the determination of molecular fields. —II. From the equation of state of a gas, Proc. R. Soc. Lond. Ser. A **106**, 463 (1924).
[41] R. E. Dinnebier and S. J. L. Billinge (eds.), *Powder Diffraction: Theory and Practice* (The Royal Society of Chemistry, Cambridge, UK, 2008).
[42] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. **68**, 314 (2013).
[43] V. Wang, N. Xu, J.-C. Liu, G. Tang, and W.-T. Geng, VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code, Comput. Phys. Commun. **267**, 108033 (2021).
[44] A. Reuss, Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle,, ZAMM - J. Appl. Math. Mech. Z. Für Angew. Math. Mech. **9**, 49 (1929).
[45] E. Wiechert, Gesetze der elastischen Nachwirkung für constante Temperatur, Ann. Phys. **286**, 335 (1893).
[46] R. Hill, The elastic behaviour of a crystalline aggregate, Proc. Phys. Soc. Sect. A **65**, 349 (1952).
[47] H. Chen and L.-X. Cai, Theoretical conversions of different hardness and tensile strength for ductile materials based on stress–strain curves, Metall. Mater. Trans. A **49**, 1090 (2018).
[48] C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli, and S. Curtarolo, High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasiharmonic Debye model, Phys. Rev. B **90**, 174107 (2014).
[49] H.-L. Shi and Z.-A. Li, Niggli reduction and Bravais lattice determination, J. Appl. Crystallogr. **55**, 204 (2022).
[50] Z. Yang, X. Wang, Y. Li, Q. Lv, C. Y.-C. Chen, and L. Shen, Efficient equivariant model for machine learning interatomic potentials, npj Comput. Mater. **11**, 49 (2025).
[51] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, Nat. Mach. Intell. **5**, 1031 (2023).
[52] See Supplemental Material at http://link.aps.org/supplemental/10.1103/8qyx-sqtn for the benchmark of MLIPs and the structure of uniaxial superhard materials.
[53] R. Gaillac, P. Pullumbi, and F.-X. Coudert, ELATE: An open-source online application for analysis and visualization of elastic tensors, J. Phys. Condens. Matter **28**, 275201 (2016).
[54] M. Popov, V. Churkin, A. Kirichenko, V. Denisov, D. Ovsyannikov, B. Kulnitskiy, I. Perezhogin, V. Aksenenkov, and V. Blank, Raman spectra and bulk modulus of nanodiamond in a size interval of 2–5 nm, Nanoscale Res. Lett. **12**, 561 (2017).
[55] J. Cui, X. Zheng, W. Bao, J.-X. Liu, F. Xu, G.-J. Zhang, and Y. Liang, Coexistence of superhardness and metal-like electrical conductivity in high-entropy dodecaboride composite with atomic-scale interlocks, Nano Lett. **23**, 9319 (2023).
[56] K. Feng, MLIPs training data, computation parameters, and corresponding results, Zenodo (2025), doi:https://doi.org/10.5281/zenodo.17420915.